

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/47, A61K 38/17	A2	(11) International Publication Number: WO 98/46757 (43) International Publication Date: 22 October 1998 (22.10.98)
(21) International Application Number: PCT/US98/07999 (22) International Filing Date: 14 April 1998 (14.04.98) (30) Priority Data: 08/843,374 15 April 1997 (15.04.97) US 09/059,487 13 April 1998 (13.04.98) US (71) Applicant: GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US). (72) Inventors: JACOBS, Kenneth; 151 Beaumont Avenue, Newton, MA 02160 (US). MCCOY, John, M.; 56 Howard Street, Reading, MA 01867 (US). LAVALLIE, Edward, R.; 113 Ann Lee Road, Harvard, MA 01451 (US). RACIE, Lisa, A.; 124 School Street, Acton, MA 01720 (US). MERBERG, David; 2 Orchard Drive, Acton, MA 01720 (US). TREACY, Maurice; 93 Walcott Road, Chestnut Hill, MA 02167 (US). SPAULDING, Vikki; 11 Meadowbank Road, Billeria, MA 01821 (US). AGOSTINO, Michael, J.; 26 Wolcott Avenue, Andover, MA 01810 (US). (74) Agent: SPRUNGER, Suzanne, A.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM (57) Abstract Polynucleotides and the proteins encoded thereby are disclosed.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

5 This application is a continuation-in-part of application Ser. No. 60/XXX,XXX (converted to a provisional application from non-provisional application Ser. No. 08/843,374), filed April 15, 1997, which is incorporated by reference herein.

FIELD OF THE INVENTION

10 The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins.

BACKGROUND OF THE INVENTION

15 Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein
20 in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization cloning techniques, have advanced the state of the art by making available large numbers of
25 DNA/amino acid sequences for proteins that are known to have biological activity by virtue of their secreted nature in the case of leader sequence cloning, or by virtue of the cell or tissue source in the case of PCR-based techniques. It is to these proteins and the polynucleotides encoding them that the present invention is directed.

30

35

```

TTTTATTAT CTCTTTGGCC CAGCCAAGCT GCCCTCACTA CAGAGACCTT GGACAAGGAT 2940
CCAGCCAGTC CCTCTCTGCC CCACAACCCT GCATTCCCAG AGGTTAGCTA TGCAGCCCAC 3000
CTAGATGAGT CTCTTCAAGA ATGGGAAATC AAGGGGTGAC AGGGAGTAAA AGGGTTATCA 3060
TCTTACTGCA AAGCCACAAG ATCAGGGCAG GGCTTTAGGA TGTCTGGAT GCTTTTTAAT 3120
AATTATGCTT CCCATCATAA CTGGGGAGAA AGGGAAGTCA GGGTTCTAGG GGTTATTCGT 3180
CCCAGGAAAT AGAAGTGAAA TTGTCTTTAT TAAGTGAAAA CTTTCCCCTT TGCCCTGCAA 3240
TG TAGCTGGG CATTCAAACG GAGGGCAAAC CGATGATCTA AACCAACCAC TTGAAAAAC 3300
CCAATGGGGA CATTGTAACC AGAGGGTCCT GGAGGTGGGG TTGATGGGTT TCCTTATCCC 3360
CAAAGTCACT CCTGTTTGT TTTGTTTTTC TTTGGGGGTT TTGTTTATTT TTGGGGCTGG 3420
CAATCCAAAA TAGAAAATCT GATCCTTTGA GGCTCTAAAG GAAAATCAGC TGCCTCTACC 3480
AACCACCCTC TATCAGCAGT GGCCAGGAA GGAGGTCAAG CATCTTCGGC CGATATTTAA 3540
ACATGGGCAG CTTCTTCAG GATGATCACC GAGGCTCCCG TGA CTCTTGAA CTCCCTACTC 3600
TCCAGAATCC AGGGGCTATA GCGATGGGGA CTGCGGAATT ACGAGGGCTG GCTGTTTTAC 3660
ACCGGTCACA TTTCTATTG GCAGTGACTG ATTCATGGGA AAGGGCTTTG AAGGAAC TAC 3720
TTCAGTGCAC ACACAAGTA CGAACCTYTC AGGCCTTTTCG AAGAACTTTC ATAATTCATG 3780
AAAGCCCAGT TYTGAAGATT CACGTATCCA TYTGGAGACC TACAGGAAGA AAGTGATTGG 3840
GTTCTCTTGG TTCTTGCTG CTTCACTGTG GATGGGAAGA GGTGACAACC TCAGTCTCCC 3900
TTTGGGACCT GTCCAAGGGT AGGCAACCAC CTTACCTTC ACACAGATTG AGGAGACACT 3960
GGACTTTTTA CCCATTTTCT TTAATYTTCA ATATTAATAT TGTGTTTACA TTGATGAGAA 4020
CAAGAGTTAA TGCCCTACCC TCTGCTGGGC TGTTTGTATT GAGTTGCAAT GTGACCAGCG 4080
AAAGCTGCAT TTAATAAATG AAAGTACAGA CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA 4140
AA 4142

```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 756 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Met Leu Gly Ile Cys Arg Gly Arg Arg Lys Phe Leu Ala Ala Ser Leu
1           5           10           15

Ser Leu Leu Cys Ile Pro Ala Ile Thr Trp Ile Tyr Leu Phe Ser Gly
20           25           30

Ser Phe Glu Asp Gly Lys Pro Val Ser Leu Ser Pro Leu Glu Ser Gln
35           40           45

Ala His Ser Pro Arg Tyr Thr Ala Ser Ser Gln Arg Glu Arg Glu Ser
50           55           60

Leu Glu Val Arg Met Arg Glu Val Glu Glu Glu Asn Arg Ala Leu Arg
65           70           75           80

Arg Gln Leu Ser Leu Ala Gln Gly Arg Ala Pro Ser His Arg Arg Gly
85           90           95

Asn His Ser Lys Thr Tyr Ser Met Glu Glu Gly Thr Gly Asp Ser Glu
100          105          110

Asn Leu Arg Ala Gly Ile Val Ala Gly Asn Ser Ser Glu Cys Gly Gln
115          120          125

Gln Pro Val Val Glu Lys Cys Glu Thr Ile His Val Ala Ile Val Cys
130          135          140

Ala Gly Tyr Asn Ala Ser Arg Asp Val Val Thr Leu Val Lys Ser Val
145          150          155          160

Leu Phe His Arg Arg Asn Pro Leu His Phe His Leu Ile Ala Asp Ser
165          170          175

Ile Ala Glu Gln Ile Leu Ala Thr Leu Phe Gln Thr Trp Met Val Pro
180          185          190

Ala Val Arg Val Asp Phe Tyr Asn Ala Asp Glu Leu Lys Ser Glu Val
195          200          205

Ser Trp Ile Pro Asn Lys His Tyr Ser Gly Ile Tyr Gly Leu Met Lys
210          215          220

Leu Val Leu Thr Lys Thr Leu Pro Ala Asn Leu Glu Arg Val Ile Val
225          230          235          240

Leu Asp Thr Asp Ile Thr Phe Ala Thr Asp Ile Ala Glu Leu Trp Ala
245          250          255

Val Phe His Lys Phe Lys Gly Gln Gln Val Leu Gly Leu Val Glu Asn
260          265          270

Gln Ser Asp Trp Tyr Leu Gly Asn Leu Trp Lys Asn His Arg Pro Trp

```

275	280	285
Pro Ala Leu Gly Arg Gly Tyr Asn Thr Gly Val Ile Leu Leu Leu Leu 290	295	300
Asp Lys Leu Arg Lys Met Lys Trp Glu Gln Met Trp Arg Leu Thr Ala 305	310	315 320
Glu Arg Glu Leu Met Gly Met Leu Ser Thr Ser Leu Ala Asp Gln Asp 325	330	335
Ile Phe Asn Ala Val Ile Lys Gln Asn Pro Phe Leu Val Tyr Gln Leu 340	345	350
Pro Cys Phe Trp Asn Val Gln Leu Ser Asp His Thr Arg Ser Glu Gln 355	360	365
Cys Tyr Arg Asp Val Ser Asp Leu Lys Val Ile His Trp Asn Ser Pro 370	375	380
Lys Lys Leu Arg Val Lys Asn Lys His Val Glu Phe Phe Arg Asn Leu 385	390	395 400
Tyr Leu Thr Phe Leu Glu Tyr Asp Gly Asn Leu Leu Arg Arg Glu Leu 405	410	415
Phe Gly Cys Pro Ser Glu Ala Asp Val Asn Ser Glu Asn Leu Gln Lys 420	425	430
Gln Leu Ser Glu Leu Asp Glu Asp Asp Leu Cys Tyr Glu Phe Arg Arg 435	440	445
Glu Arg Phe Thr Val His Arg Thr His Leu Tyr Phe Leu His Tyr Glu 450	455	460
Tyr Glu Pro Ala Ala Asp Ser Thr Asp Val Thr Leu Val Ala Gln Leu 465	470	475 480
Ser Met Asp Arg Leu Gln Met Leu Glu Ala Ile Cys Lys His Trp Glu 485	490	495
Gly Pro Ile Ser Leu Ala Leu Tyr Leu Ser Asp Ala Glu Ala Gln Gln 500	505	510
Phe Leu Arg Tyr Ala Gln Gly Ser Glu Val Leu Met Ser Arg His Asn 515	520	525
Val Gly Tyr His Ile Val Tyr Lys Glu Gly Gln Phe Tyr Pro Val Asn 530	535	540
Leu Leu Arg Asn Val Ala Met Lys His Ile Ser Thr Pro Tyr Met Phe 545	550	555 560
Leu Ser Asp Ile Asp Phe Leu Pro Met Tyr Gly Leu Tyr Glu Tyr Leu 565	570	575

Arg Lys Ser Val Ile Gln Leu Asp Leu Ala Asn Thr Lys Lys Ala Met
 580 585 590
 Ile Val Pro Ala Phe Glu Thr Leu Arg Tyr Arg Leu Ser Phe Pro Lys
 595 600 605
 Ser Lys Ala Glu Leu Leu Ser Met Leu Asp Met Gly Thr Leu Phe Thr
 610 615 620
 Phe Arg Tyr His Val Trp Thr Lys Gly His Ala Pro Thr Asn Phe Ala
 625 630 635 640
 Lys Trp Arg Thr Ala Thr Thr Pro Tyr Arg Val Glu Trp Glu Ala Asp
 645 650 655
 Phe Glu Pro Tyr Val Val Val Arg Arg Asp Cys Pro Glu Tyr Asp Arg
 660 665 670
 Arg Phe Val Gly Phe Gly Trp Asn Lys Val Ala His Ile Met Glu Leu
 675 680 685
 Asp Val Gln Glu Tyr Glu Phe Ile Val Leu Pro Asn Ala Tyr Met Ile
 690 695 700
 His Met Pro His Ala Pro Ser Phe Asp Ile Thr Lys Phe Arg Ser Asn
 705 710 715 720
 Lys Gln Tyr Arg Ile Cys Leu Lys Thr Leu Lys Glu Glu Phe Gln Gln
 725 730 735
 Asp Met Ser Arg Arg Tyr Gly Phe Ala Ala Leu Lys Tyr Leu Thr Ala
 740 745 750
 Glu Asn Asn Ser
 755

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1435 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGCACCGGTG GTCGGCTGTT GGGTGTGGAG TTTCCCAGCG CCCCTCGGGT CCGACCCTTT 60
 GAGCGTTC TG CTCCGGCGCC AGCCTACCTC GCTCCTCGGC GCCATGACCA CAACCACCAC 120

AAW74724
ID AAW74724 standard; Protein; 756 AA.
XX
AC AAW74724;
XX
DT 15-FEB-1999 (first entry)
XX
DE Human secreted protein fa252_8.
XX
KW Secreted protein; human; fa252_8
XX
OS Homo sapiens.

PN WO9846757-A2.
XX
PD 22-OCT-1998.
XX
PF 14-APR-1998; 98WO-US07999.
XX
PR 13-APR-1998; 98US-0059487.
PR 15-APR-1997; 97US-0843374.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-568731/48.
DR N-PSDB; AAV62750.
XX
PT New polynucleotide(s) encoding secreted human proteins - are derived
PT from, e.g. human foetal brain or foetal kidney cDNA libraries,
PT potentially useful as, e.g. vaccines or thrombolytic agents
XX
SQ Sequence 756 AA;

Query Match 3.9%; Score 27; DB 19; Length 756;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 GLVENQSDWYLGNLWKNHRPWPALGRG 227
|||||
Db 268 GLVENQSDWYLGNLWKNHRPWPALGRG 294